Insect Molecular Biology (2005) **14**(3), 309–318

Molecular cloning and characterization of two digestive serine proteases from the Hessian fly, *Mayetiola destructor*

O. Mittapalli*, J. J. Stuart* and R. H. Shukle†

**Department of Entomology, Purdue University, West Lafayette, IN, USA,* †*USDA-ARS, Department of Entomology, Purdue University, West Lafayette, IN, USA*

Abstract

Full-length cDNA and genomic sequences for two genes (designated *mdesprot-I* and *mdesprot-II*) encoding digestive serine proteases in Hessian fly, *Mayetiola destructor*, have been cloned and characterized. The deduced amino acid sequences revealed similarity with trypsin-like digestive serine proteases from other Dipterans. Both *mdesprot-I* and *mdesprot-II* encoded proteins with secretion signal peptides at the N-terminals, indicating the proteins are secreted proteases that should function as midgut digestive proteases. A cytological analysis with fluorescent *in situ* hybridization revealed the cytological localization of *mdesprot-I* and *mdesprot-II* on the long arm of Autosome 2. Results are discussed in the context of the efficacy of potential protease inhibitors to develop Hessian fly resistant wheat through genetic engineering approaches.

Keywords: Hessian fly, *Mayetiola destructor*, midgut, digestive serine protease, wheat.

Introduction

The Hessian fly, *Mayetiola destructor*, is a destructive pest of wheat, *Triticum aestivum*, and poses a serious concern in several wheat growing regions of the world. Other hosts of this pest when wheat is not available include rye, barley, triticale and wild grasses such as *Agropyron repens*, *Hordeum pusillum*, *Elmus virginicus*, and *Aegilops* sp. (Jones, 1939; Shukle, 2003).

Resistant genes controlling larval antibiosis deployed into wheat are the most effective method of control (Gallun, 1977). However, due to the deployment of resistant cultivars, the pest has evolved biotypically to overcome the resistance offered by formally resistant wheat. In the United States, 31 genes for resistance, designated H_1 to H_{31}, which are effective against the Hessian fly have been identified in *Triticum* sp. and *Secale cereale* (L.). (Poaceae) (Cox & Hatchett, 1994; Ohm *et al*., 1997; Martin-Sanchez *et al*., 2003; Williams *et al*., 2003). However, the emergence of more virulent biotypes to many of the known genes, remains a threat for long-term protection (Ratcliffe *et al*., 2000). Thus, there is a need for a better understanding of the interaction between the Hessian fly and its host plant that would enhance the success of continued protection of wheat from this pest.

The 1st-instars are the most destructive stage of the life cycle. Soon after hatching, larvae crawl into the region between the first and second leaf sheaths and commence to feed near the crown on seedling plants, or just above the nodes in the case of jointed wheat (Shukle, 2003). Systemic symptoms of damage to wheat seedlings include stunting and development of a dark green colour and can ultimately lead to death of seedling plants (Cartwright *et al*., 1959). Even a single larva per plant is sufficient to evoke symptoms, which are thought to be caused by the secretion of unknown substances from the salivary glands (Byers & Gallun, 1972; Stuart & Hatchett, 1988). These secretions are believed to contain enzymes that increase permeability of cell membranes, weaken cell walls and inhibit plant growth and development (Byers & Gallun, 1972).

There have been no reports of cloned genes encoding digestive proteases in the Hessian fly larvae. However, Shukle *et al*. (1985) observed that the major proteolytic activity in the guts of larvae was chymotrpysin-like. The nature of this proteolytic activity seemed to be consistent in expression in all larval stages. Our interest in the digestive

doi: 10.1111/j.1365-2583.2005.00561.x

Received 18 August 2004; accepted after revision 3 January 2005. Correspondence: Dr Richard H. Shukle, USDA-ARS, Department of Entomology, 901 W. State St, Purdue University, West Lafayette, IN 47906, USA. Tel.: +1 765 4946351; fax: +1 765 4961219; e-mail: shukle@purdue.edu

enzyme profile of Hessian fly larvae stems from their possible involvement in damage caused to the host plant by feeding larvae. Additionally, digestive proteases expressed in the larval gut may serve as sites of action for chemical defences to be conferred on plants that could impede the development of virulence to native resistance genes in wheat (Shukle *et al.*, 1985). The Bowman-Birk inhibitor (BBI) from soybean seeds (*Glycine max*) has two inhibitory domains and the ability to inhibit both trypsin- and chymotrypsin-like proteolytic activities occurring in the guts of phytophagous insects (Rahbe *et al.*, 2003). The evidence that protease inhibitors can serve as insect defences includes retardation of growth of several insect species when protease inhibitors from plants were incorporated into artificial diets (Gatehouse *et al.*, 1979). Hence, it seems prudent to develop such alternative control strategies which, along with native resistant genes, would ensure resistance in future wheat cultivars.

In this paper, we report the cloning and characterization of two genes and related cDNAs that encode trypsin-like digestive serine proteases; this is intended as a basis for future studies on the Hessian fly midgut as an interface with its host plant.

Results

Characterization of mdesprot-I *and* mdesprot-II *cDNA and genomic sequences*

Full-length cDNA sequences for two digestive serine proteases were recovered using RNA extracted from 1st-instars of the Hessian fly by the 5′- and 3′ RACE procedure using gene specific primers anchored in the partial sequences obtained via reverse transcription polymerase chain reaction (RT-PCR) with degenerate primers. Degenerate primers designed from chymotrypsin-like proteases of several Dipterans amplified the partial sequence for *mdesprot-I*, while the primers designed from trypsin-like proteases amplified the partial sequence of *mdesprot-II (*see Fig. 1). The cDNA sequence for *mdesprot-I* was 807 bp, and the genomic sequence for the coding region was 1137 bp. For *mdesprot-II* the cDNA was 822 bp, and the genomic sequence for the coding region was 1151 bp. Three introns spanned the genomic sequences of both *mdesprot-I* and *mdesprot-II* with introns two and three in close proximity and located towards the 3′ region of the genes (Fig. 1). No introns were identified within the 5′ or 3′ untranslated regions of either *mdesprot-I* or *mdesprot-II*. The full-length cDNA and genomic sequences were in accordance in regard to their sizes when compared to digestive serine proteases of various other insects including other Dipterans. The nucleotide sequences of both the proteases were submitted to GenBank and have the following accession numbers: *mdesprot-I*: AY664493; *mdesprot-II*: AY664494.

Cytological localization of mdesprot-I *and* mdesprot-II *genes*

To determine whether both the cloned proteases of the Hessian fly were encoded in its genome and to assess their localization, a cytological analysis was conducted with fluorescence *in situ* hybridization (FISH). Two bacterial artificial chromosomes (BAC) clones, 45B9 and hf9K1, that contained both the proteases were labelled, respectively, with biotin and digoxigenin and hybridized *in situ* to salivary polytene chromosomes. The BAC clones used in this study have been described elsewhere (Behura *et al.*, 2004). An overlapping hybridization signal (indicated as a yellow signal) was observed (Fig. 2) on both chromatids of the long arm of Polytene Chromosome 2 (i.e. Autosome A2, Stuart & Hatchett, 1988). However, with the BAC clone hf9K1 there were additional signals (red signals) observed near the centromeres. This suggests the presence of repetitive elements such as *mariner* transposable elements in this clone. Both proteases were located within a region of the long arm of Chromosome A2 where previous studies (Chen *et al.*, 2004; Liu *et al.*, 2004) and on-going studies have revealed a zone of low-recombination and where a number of other genes have been cytologically positioned (Behura *et al.*, 2004).

Expression of mdesprot-I *and* mdesprot-II *mRNA in larval tissues and during development*

RT-PCR with different numbers of amplification cycles using RNA extracted from midguts of late 1st-instars and early 2nd-instars, indicated that for both the Hessian fly proteases (*mdesprot-I* and *mdesprot-II*) mRNA was expressed in the midgut tissue of the larvae. However, RT-PCR using RNA extracted from the salivary glands or the remaining carcasses (includes fat body, neural tissue, muscle, etc.) of the same larvae after removal of the midgut, did not result in the production of an amplicon for either protease (Fig. 3). RT-PCR with primers designed to a Hessian fly actin gene produced an amplicon of the expected size, with RNA from midgut tissue, salivary glands, and larval carcasses. RT-PCR without reverse transcriptase did not result in the production of an amplicon with RNA from the tissues.

Analysis of *mdesprot-I* and *mdesprot-II* mRNA in larvae, pupae, and adults by RT-PCR with different numbers of amplification cycles indicated that mRNA for *mdesprot-I* was expressed in 1st-instars and early 2nd-instars but not in late 2nd-instars, 3rd-instars, pupae, or adults (Fig. 4). Similarly, RT-PCR indicated that mRNA for *mdesprot-II* was expressed in 1st-instars and early 2nd-instars. However, RT-PCR indicated that mRNA for *mdesprot-II* was also expressed in non-feeding larval stages (i.e. late 2nd-instars and 3rd-instars; Gallun & Langston, 1963) as well as in pupae but not in adults (Fig. 4). While not quantitative PCR, the present results indicate the amplicon for *mdesprot-II* has not been taken to saturation with 30 amplification

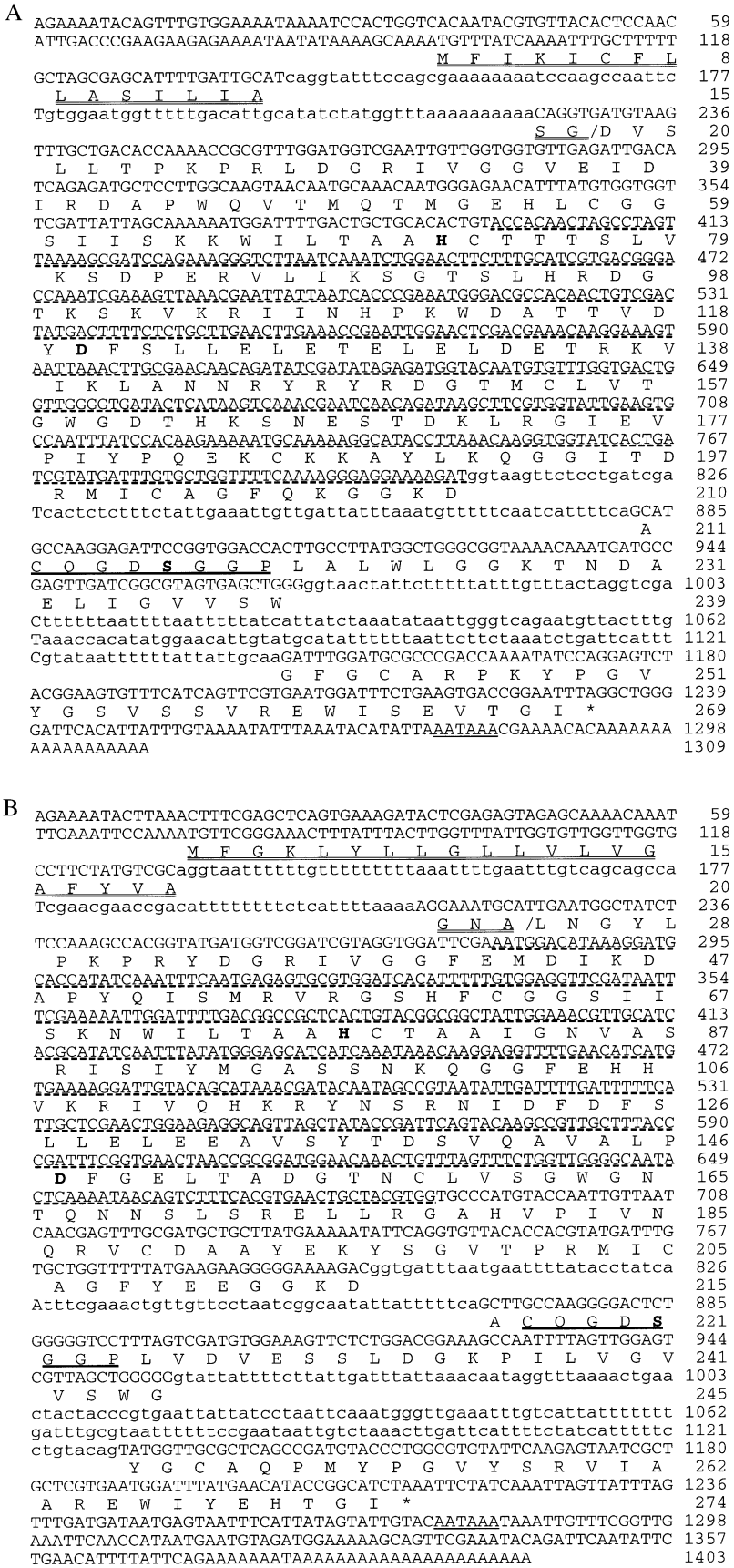

Figure 1. Nucleotide sequence and predicted amino acid sequence of Hessian fly *mdesprot-I* and *mdesprot-II* genes. (A) *mdesprot-I*, 5′- and 3′ untranslated regions and exon sequences are represented in upper case nucleotides. Intron sequences are indicated in lower case nucleotides. The stop codon (TAG) is shown as an asterix. An AATAAA sequence matching the consensus polyadenylation signal in other species is underlined. The predicted signal peptide sequence is double underlined and the most likely cleavage site between amino acid residues glycine (G17) and aspartic acid (D18) is indicated by a solidus (/). Amino acid residues in the active site triad histidine (H72), aspartic acid (D120) and serine (S216) are shown in bold. An active pocket sequence that should confer trypsin-like specificity is underlined in bold. The partial sequence obtained with degenerate primers designed to chymotrypsin-like proteases is underlined with a dashed line. (B) *mdesprot-II*, 5′- and 3′ untranslated regions and exon sequences are represented in uppercase nucleotides. Intron sequences are indicated in lowercase nucleotides. The stop codon (TAA) is shown as an asterix. An AATAAA sequence matching the consensus polyadenylation signal in other species is underlined. The predicted signal peptide sequence is double underlined and the most likely cleavage site between amino acid residues A23 and L24 is indicated by a solidus (/). Amino acid residues in the active site triad H77, D147 and S221 are shown in bold. An active pocket sequence that should confer trypsin-like specificity is underlined in bold. The partial sequence obtained with degenerate primers designed to trypsin-like proteases is underlined with a dashed line.

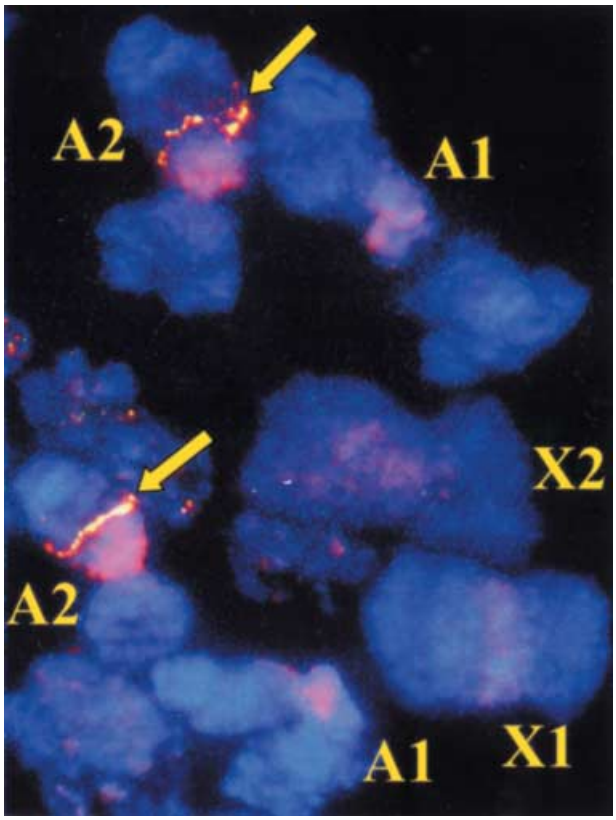

Figure 2. Results of fluorescent *in situ* hybridization (FISH) with *mdesprot-I* and *mdesprot-II* to Hessian fly Polytene chromosomes. Two BAC clones, hf9K1 (digoxigenin labelled) and 45B9 (biotin labelled), both containing the *mdesprot-I* and *mdesprot-II* genes were used as the probes. An overlap of the hybridization signals (yellow fluorescence) from the two clones was observed on both chromatids of the long arm of Autosome 2 (A2). The nucleolus (N) on Autosome 1 (A1) is indicated. Red hybridization signals near the centromeric regions with BAC hf9K1 suggest the presence of a repetitive element in this clone. X1 and X2 represent the sex chromosomes of Hessian fly.

cycles using equal amounts of RNA from late 2nd-instars, 3rd-instars, and pupae (Fig. 4).

Characterization of mdesprot-I *and* mdesprot-II *proteins*

A signal peptide was represented in the first 17 amino acid residues of *mdesprot-I* and in the first 23 amino acid residues of *mdesprot-II* (Fig. 1). Additionally, both proteases correctly positioned the histidine, aspartic acid, and serine members of the catalytic triad and contained an active pocket sequence containing the catalytic serine residue, as found in most trypsins from Dipterans as well as non-dipterous insects (Fig. 1). The calculated molecular weight for *mdesprot-I* was 29.5 kDa; while for *mdesprot-II* the molecular weight was 29.8 kDa. Alignment of the deduced amino acid sequence of *mdesprot-I* revealed 44–47% similarity with trypsins of *Culex pipens pallens*, *Anopheles gambiae* and *Drosophila melanogaster*, while the deduced amino acid sequence of *mdesprot-II* displayed 53%, 54% and 57% similarity, respectively, with trypsins of *D. melanogaster, A. gambiae* and *C. pipens pallens*.

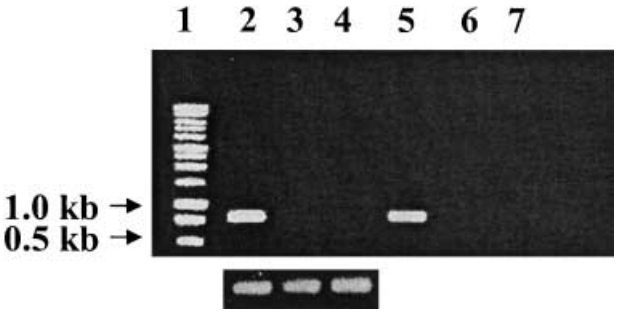

Figure 3. Agarose gel analysis of RT-PCR product obtained with 30 amplification cycles using primers designed to the ends of the coding regions of the cDNAs for *mdesprot-I* and *mdesprot-II*, respectively, and RNA extracted from Hessian fly larval midgut tissue, salivary glands and larval carcasses after removal of the midgut and salivary glands. A PCR product was detected for *mdesprot-I* with RNA from midgut tissue (lane 2). No PCR product for *mdesprot-I* was detected with RNA from salivary glands (lane 3) or larval carcasses (lane 4). For *mdesprot-II* a PCR product was detected with RNA from midgut tissue (lane 5). No PCR product for *mdesprot-II* was detected with RNA from salivary glands (lane 6) or larval carcasses (lane 7). Each PCR (5 μl of a 50 μl reaction volume) was loaded per lane. A 1 kb DNA ladder (Promega) was run in lane 1. RT-PCR product obtained with primers designed to a Hessian fly actin gene and RNA extracted from midgut tissue (lane 2), salivary glands (lane 3), and remaining larval carcasses (lane 4) is shown as an insert.

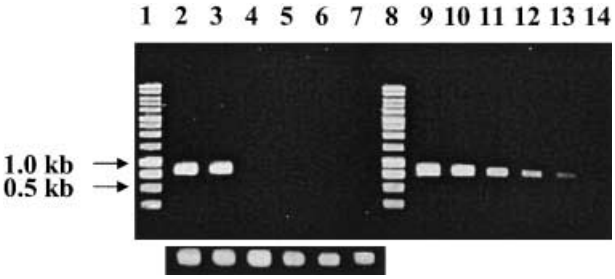

Figure 4. Agarose gel analysis of RT-PCR product obtained with 30 amplification cycles using primers designed to the ends of the coding regions of the cDNAs for *mdesprot-I* and *mdesprot-II*, respectively, and RNA extracted from Hessian fly larvae, pupae, and adults. A PCR product was detected for *mdesprot-I* with RNA from 1st-instar (lane 2) and early 2nd-instar larvae (lane 3). No PCR product was detected for *mdesprot-I* with RNA from late 2nd-instar larvae (lane 4), 3rd-instar larvae (lane 5), pupae (lane 6), or adults (lane 7). For *mdesprot-II*, a PCR product was detected with RNA from 1st-instar larvae (lane 9), early 2nd-instar larvae (lane 10), late 2nd-instar larvae (lane 11), 3rd-instar larvae (lane 12), and pupae (lane 13). No PCR product was detected for *mdesprot-II* with RNA from adults (lane 14). 5 μl of each PCR (50 μl reaction volume) was loaded per lane. A 1 kb DNA ladder (Promega) was run in lane 1 and lane 8. RT-PCR product obtained with primers designed to a Hessian fly actin gene and RNA extracted from 1st-instars (lane 2), early 2nd-instars (lane 3), late 2nd-instars (lane 4), 3rd-instars (lane 5), pupae (lane 6) and adults (lane 7) is shown as an inset.

Relationship of mdesprot-I *and* mdesprot-II *to other insect midgut serine proteases*

To reveal the relationship of the Hessian fly midgut proteases *mdesprot-I* and *mdesprot-II* to other insect midgut serine proteases, a distance/neighbor-joining tree was constructed, with trypsin-like and chymotrypsin-like proteases

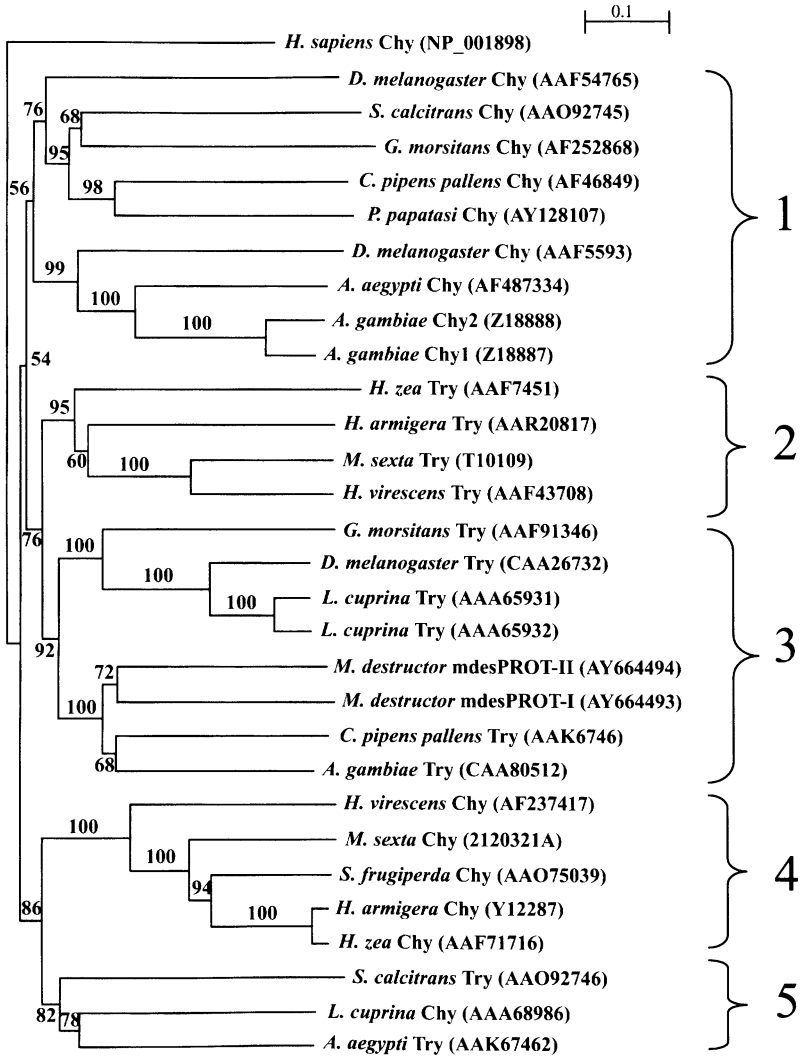

Figure 5. A phylogenetic tree based on the amino acid sequences for the midgut digestive serine proteases *mdesprot-I* and *mdesprot-II* from Hessian fly and digestive serine proteases from other insects. A *Homo sapiens* pancreatic chymotrypsin was included as the outgroup. The topology and branch lengths of the phylogram were produced by the distance/neighbor-joining method. The percentage of 1000 bootstrap replications supporting each node is shown. Accession numbers for the proteases are shown in parentheses.

from a number of Dipterans and Lepidoptera, and using a pancreatic chymotrypsin from *Homo sapiens* as the outgroup. The tree revealed five major clades with bootstrap support: (1) chymotrypsin-like proteases from all but one of the Dipterans included in the analysis; (2) trypsin-like proteases from all of the Lepidoptera included in the analysis; (3) trypsin-like proteases from all but two of the Dipterans; (4) chymotrypsin-like proteases from the Lepidoptera; (5) trypsin-like proteases from *Stomoxys calcitrans* and *Aedes aegypti* and a chymotrypsin-like protease from *Lucilia cuprina* (Fig. 5). The midgut proteases *mdesprot-I* and *mdesprot-II* from Hessian fly grouped within the third clade (trypsin-like proteases from the Dipterans) and specifically in a subclade with trypsins from *A. gambiae* and *C. pipens pallens*.

Discussion

To date, the most effective means of control for the Hessian fly has been via genetic resistance in the host plant (El Bouhssini *et al.*, 2001). This mode of resistance is expressed by larval antibiosis and is governed predominantly by single genes that are completely or partially dominant (Zantoko & Shukle, 1997; El Bouhssini *et al.*, 1998). Hatchett & Gallun (1970) stated the occurrence of a close genetic interaction between wheat and the Hessian fly; however, the chemical basis of resistance in wheat has yet to be unravelled.

Serine proteases have been well known to participate in various physiological and cellular functions, of which their role in digestion is one of the most extensively studied aspects (Krem & Cera, 2001). Chymotrypsins and trypsins are representatives of one of the largest serine protease families, wherein, the former hydrolyses peptide bonds on the carboxyl terminus of aromatic amino acids (tryptophan, tyrosine, and phenylalanine), and the latter hydrolyses peptide bonds involving amino acids with positively charged side chains such as arginine and lysine (Kraut, 1977). Like many enzymes, serine proteases are synthesized as inactive

precursors (zymogens) that consist of a signal peptide and pro-enzyme sequence at the amino-terminus. Upon reaching their sites of action, these accessory regions (signal and pro-enzyme sequences) are clipped off by other enzymes, resulting in the production of active enzymes.

The deduced amino acid sequences of the two Hessian fly serine proteases obtained in this study revealed significant similarities with trypsin-like serine proteases from other Dipterans, specifically from *A. gambiae* and *C. pipens pallens.* Additionally, both proteases correctly positioned conserved amino acid residues of the serine protease active site triad as well as residues within an active site pocket containing the serine residue that help to confer trypsin-like specificity to the enzymes (Fig. 1). Within *mdesprot-I*, the amino acid residues in the active site triad were His72, Asp120, and Ser216, with the serine residue positioned in the active pocket sequence Cys121, Gln213, Gly214, Asp215, Ser216, Gly217, Gly218 and Pro219, characteristic of trypsins from most insects (Zhu *et al.*, 2000). Within *mdesprot-II*, the amino residues in the active site triad were His77, Asp147 and Ser221. The active pocket sequence was Cys217, Gln218, Gly219, Asp220, Ser221, Gly222, Gly223 and Pro224. A phylogenetic tree (Fig. 5) revealed both *mdesprot-I* and *mdesprot-II* grouped within the clade containing trypsin-like digestive proteases from other Dipterans and within a subclade that contained trypsin-like proteases from *A. gambiae* and *C. pipens pallens*. The topology of the clades contained in the phylogenetic tree was well supported and congruent with the expected relationships for proteases from the Dipterans and Lepidoptera. This further supports the hypothesis that *mdesprot-I* and *mdesprot-II* are trypsin-like digestive proteases of Hessian fly.

Results obtained with fluorescence *in situ* hybridization (FISH) indicated the genes for both proteases were localized to the same region of the long arm of Polytene Chromosome 2, Autosome A2. The region of Chromosome A2 that *mdesprot-I* and *mdesprot-II* localized to is the region where a family of salivary gland transcripts *SSGP-10A1* (Chen *et al.*, 2004), the salivary gland gene *SSGP-12A1* (Liu *et al.*, 2004), and two avirulence (*avr*) genes *vH3* and *vH5* (Behura *et al.*, 2004) have been positioned cytologically. The presence of additional signals near the centromeric regions of BAC clone hf9K1 suggests this clone contains repetitive elements. Indeed, recent work has demonstrated the clone does contain the *mariner* transposon *desmar1* (Russell & Shukle, 1997).

To date, there have been no reports of cloning and characterization of digestive serine proteases from the Hessian fly. However, Shukle *et al.* (1985) reported chymotrypsin-like serine protease activity in the midguts of feeding larvae. They indicated the proteolytic activity was associated with gut lumen contents, and the nature of proteolytic activity was consistent in all feeding larval stages.

Evidence that both of the cloned proteases of Hessian fly in the current study are secreted into the lumen of the midgut is the presence of secretion signal peptide sequences at the N-terminus in the deduced amino acid sequence for both proteases (Fig. 1). The length of the putative signal peptides associated with *mdesprot-I* and *mdesprot-II* were in agreement with those reported for trypsins from other Dipterans and non-dipterous insects (Davis *et al.*, 1985; Kalhok *et al.*, 1993; Kollien *et al.*, 2004). Additionally, RT-PCR with RNA extracted from midguts, salivary glands and the remaining carcasses of larvae, indicated that for *mdesprot-I* and *mdesprot-II*, mRNA was detected in midgut tissue but not in the other tissues assayed (shown in Fig. 3). These results support the hypothesis that both proteases are targeted to the lumen of the midgut, are midgut specific in expression, and should perform extra-cellular digestion of proteins within the gut lumen of feeding Hessian fly larvae.

Both *mdesprot-I* and *mdesprot-II* were expressed in feeding larvae (1st-instars and early 2nd-instars). However, *mdesprot-II* was also expressed in non-feeding larvae (late 2nd-instars and 3rd-instars) and pupae. While caution needs to be exercised in interpreting the results of RT-PCR, the expression of mRNA for *mdesprot-II* appeared to be decreasing in non-feeding larvae and pupae (Fig. 4). These results suggest *mdesprot-II* should have a function during development in the midguts of non-feeding larvae and pupae. To determine whether this function is the continued digestion of proteins imbibed during feeding, or an additional function, would require further study.

The expression of the proteases *mdesprot-I* and *mdesprot-II* in the midguts of feeding larvae indicates they could be targets for protease inhibitors as transgenes for Hessian fly resistance in wheat. While various studies have supported the function of protease inhibitors as antiherbivore defences (Shirai, 2004; Zavala *et al.*, 2004), other studies have revealed that phytophagous insects can adapt to the presence of protease inhibitors in their diet by various mechanisms, including up-regulation of protease genes or *de novo* synthesis of proteases that have modifications in key amino acids rendering them resistant to the inhibitors (Gruden *et al.*, 2003; Volpicella *et al.*, 2003; Brown *et al.*, 2004; Lopes *et al.*, 2004).

The most practical method of controlling the Hessian fly has been the deployment of resistant cultivars. However, the development of Hessian fly genotypes that can survive on formally resistant wheat is a threat to the durable protection of wheat from this pest. Within the south-east United States, Hessian fly populations have developed virulence to resistance genes *H3*, *H5*, *H6*, and *H7H8* in response to their deployment in wheat cultivars (Ratcliffe *et al.*, 2000). Our interests are directed toward the identification of potential transgenes for Hessian fly resistance that should have different modes of action from native resistance genes in wheat and can be pyramided with native genes for more

durable resistance. Previous work (Shukle *et al*., 1985) has indicated that the digestive serine proteases in the lumen of the Hessian fly larval midgut can be effectively inhibited by the Bowman-Birk inhibitors (BBIs). Furthermore, this group of inhibitors has been documented to be effective against fluid-sucking insects such as aphids (Rahbe *et al*., 2003). Thus, a BBI gene may prove an effective transgene for Hessian fly resistance to combine with one or more of the currently undeployed native resistance genes.

Recent advances in molecular technology applied to the Hessian fly would enhance the probability of developing such alternative strategies. This includes the possible use of Wheat Streak Mosaic Virus (WSMV) as a transient expression vector (Choi *et al*., 2000) for potential transgenes, such as the Bowman-Birk Inhibitor (BBI), for resistance in wheat as well as the expression of insect genes involved in the compatible and incompatible interactions. Additionally, RNA interference (RNAi) appears to provide an approach to revealing the function of genes expressed in 1st-instars of the Hessian fly (Yoshiyama & Shukle, 2004). Preliminary RNAi experiments, based on a 500-bp fragment of *mdesprot-I*, have been performed. Initial results from these experiments suggested that expression of *mdesprot-I* is required for normal development of 1st-instars on the host plant, as larvae hatching from embryos injected with double-stranded RNA (dsRNA) showed significant delays in development compared to buffer-injected controls or controls injected with a dsRNA control not expected to silence/suppress Hessian fly genes (O. Mittapalli & R.H. Shukle, unpublished data).

The recovery of additional serine proteases and other midgut expressed genes in 1st-instars and 2nd-instars based on midgut Expressed Sequence Tags (ESTs) has been undertaken. These studies should give an understanding of gene expression in the midgut of the Hessian fly larvae during compatible and perhaps incompatible interactions with wheat. Such knowledge can reveal additional target sites for potential transgenes for resistance to Hessian fly that could be pyramided with native resistance genes in wheat.

Experimental procedures

Experimental insect

Hessian flies used in this study were reared on susceptible wheat seedlings (cv. 'Newton') in growth chambers at 20 °C with a 12 h photoperiod. Larval instars, pupae, and adults of laboratory Biotype Great Plains (GP) were used for extraction of RNA. Sixteen biotypes of Hessian fly (designated GP and A to O) have been identified (Ratcliffe *et al*., 1994). These biotypes are distinguished only by their ability to survive on and stunt (virulent) or inability to survive on and stunt (avirulent) wheats carrying specific genes for resistance (i.e. *H3*, *H5*, *H6*, *H7H8*). Biotype GP is defined as being aviruent with respect to each of the four resistance genes used to distinguish the biotypes. The laboratory culture of GP was established from a field collection made in Ellis County, KA. Larvae (1st-, 2nd-, and 3rd-instars) and pupae were collected by dissecting the crown portions of infested wheat seedlings immersed in water and immediately flash-frozen in liquid nitrogen. Adults were collected after emergence, cold anaesthetized, and flash-frozen in liquid nitrogen. Larvae, pupae, and adults were stored at −70 °C until RNA was isolated.

Salivary gland, midgut dissections and RNA isolation

About three hundred midguts were dissected from late 1st-instar larvae (five-days old) and early 2nd-instar larvae (eight-days old) immersed in ice-cold Schneider's insect medium (Sigma-Aldrich, St Louis, MO). Midguts were removed by first pinching off the posterior tip of a larva and then gently compressing the body, commencing from the anterior end, resulting in the expulsion of the gut (Grover *et al*., 1988). Salivary glands from late 1st-instars and early 2nd-instars were dissected in Schneider's insect medium by grasping the posterior end of a larva with a pair of forceps while a second pair of forceps was used to grasp and pull away the anterior tip of the larva with salivary glands attached to the mouthparts. The salivary glands were then removed from connecting tissues. Midguts were then expelled from the body of the larva (*vide supra*). Midguts, salivary glands, and the remaining larval carcasses after removal of midguts and salivary glands, were collected in 100 μl of ice-cold Schneider's contained in a 1.5 ml Eppendorf microcentrifuge tube. Immediately following collection, midguts, salivary glands and carcasses were flash-frozen in liquid nitrogen and stored at −70 °C until RNA was isolated. Total RNA was extracted from whole larvae and isolated tissues using the RNAqueous®-4PCR kit from Ambion following the manufacturer's protocol.

Recovery of protease sequences by reverse transcription-polymerase chain reaction (RT-PCR)

RT-PCR to initially amplify serine protease sequences from Hessian fly was performed with nested, degenerate primers based on conserved regions of trypsin- and chymotrypsin-like serine midgut digestive proteases from several Dipterans (i.e. *D. melanogaster*, *A. gambiae*, *Ae. aegypti*, *C. pipens*, and *Glossina morsitans*). RT-PCRs were carried out using the SuperScript™ One Step RT-PCR System with Platinum® *Taq* DNA polymerase (Invitrogen, Carlsbad, CA). For each RT-PCR, 1.0 μg of total RNA either from whole larvae or midguts was used. PCR cycling conditions were: 50 °C for 30 min, 94 °C for 2 min, one cycle; 94 °C for 30 s, 52 °C–55 °C for 30 s, 72 °C for 1 min, 35 cycles; 72 °C for 5 min, one cycle; hold at 4 °C.

Degenerate PCR primers designed from trypsin-like proteases were TryF, 5′-CGCATCGTDGGYGGATTCSAG-3′; TryR, 5′-CCYTGGCA-RGMMTCCTTRCC-3′; TryF-nested, 5′-CCSYAYCARGTKTCSCTGC-3′; TryR-nested, 5′-GGRCCWCCRGAGTCWCCYTGGCA-3′. Degenerate PCR primers designed from chymotrypsin-like proteases were ChyF, 5′-GGWCAYAACTGTGGWKC-3′; ChyR, 5′-GGGCCKCCKGAATCACCYTTG-3′; ChyF-nested, 5′-TGGGT-WCTRACRGCMGCACAYTGT-3′; ChyR-nested, 5′-TCACCRTT-GCAYGCWTACATYCTMAC-3′, where, R = A or G; Y = C or T; D = A, G or T; K = G or T; M = A or C; S = C or G; W = A or T.

Rapid amplification of cDNA ends (RACE)

Rapid amplification of cDNA ends (5′- and 3′ RACE) amplicons were obtained for both proteases using the GeneRacer™ kit from Invitrogen following the manufacturer's specifications. Gene specific

nested primers for 5′- and 3′ RACE were designed to the RT-PCR amplicons obtained with degenerate primers and revealed by sequence analysis to represent serine protease sequences.

Typsin-like 5′ RACE primers were Try-5′RACE, 5′-TGAATATTTTT-CATAAGCAGCATCGCAAAC-3′; Try-5′RACE-nested, 5′-CCGCG-GTTAGTTCACCGAAATCGGGTAA-3′. Chymotrypsin-like 5′ RACE primers were Chy-5′RACE, 5′-CGTCGAGTTCCAATTCGGTT-TCAAG-3′; Chy-5′RACE-nested, 5′-TGGTCCCGTCACGAT-GCAAAGAAGT-3′. Trypsin-like 3′ RACE primers were Try-3′RACE, 5′-ACAGCATAAACGATACAATAGCCGTAATAT-3′; Try-3′RACE-nested, 5′-GAGGCAGTTAGCTATACCGATTCAGTACA-3′. Chymotrypsin-like 3′ RACE primers were Chy-3′RACE, 5′-ACTTCTTTGCATCGT-GACGGGAACA-3′; Chy-3′RACE-nested, 5′-CTTGAAACCGAA-TTGGAACTCGACG-3′.

Recovery of full-length cDNA clones as well as genomic clones for the coding regions and the 3′ and 5′ untranslated regions of the Hessian fly proteases

Full-length cDNA clones for the coding regions of Hessian fly proteases were recovered by RT-PCR using RNA isolated from 1st-instar larvae of Biotype GP. Genomic clones for the coding regions as well as the 5′- and 3′ untranslated regions of the proteases were recovered by allele-specific PCR using DNA isolated from Biotype GP adults as the template. PCR primers designed to the ends of the coding regions used to amplify the cDNA and genomic sequences were Prot-IF, 5′-ATGTTTATCAAAATTTGCTTTTTGC-3′; Prot-IR, 5′-CTAAATTCCGGTCACTTCAGAAATC-3′; Prot-IIF, 5′-ATGTTCGGGAAACTTTATTTACTTG-3′; Prot-IIR, 5′-TTAGAT-GCCGGTATGTTCATA-3′.

PCR primers designed to amplify the 5′-and 3′ untranslated regions were Prot-I 5′UTR F, 5′-AGAAAATACAGTTTGTGGAAAAT-3′; Prot-I 5′UTR R, 5′-CAAATTTTGATAAACATTTTGCT-3′; Prot-I 3′ UTR F, 5′-ATTTCTGAAGTGACCGGAATTTAG-3′; Prot-I 3′ UTR R, 5′-GTGTTTTCGTTTATTTAATATGTATT-3′; Prot-II 5′ UTR F, 5′-AGAAAATACTTAAACTTTCGAGCT-3′; Prot-II 5′ UTR R, 5′-AATAAAGTTTCCCGAACATTTTGGA-3′; Prot-II 3′ UTR F, 5′-ATTTATGAACATACCGGCATCTAA-3′; Prot-II 3′ UTR R, 5′-CTGAATAAAATGTTCAGAATATTG-3′.

PCR was performed in a 50-μl reaction volume using Invitrogen Platinum *Taq* DNA Polymerase High Fidelity, according to the manufacturer's instructions. Temperature cycling was according to the manufacturer's protocol. Annealing temperatures for the *mdesprot-I* and *mdesprot-II* primers designed to the ends of the coding regions were 59 °C and 55 °C, respectively. Annealing temperature for the primers designed to the ends of the 5′- and 3′ untranslated regions of *mdesprot-I* and *mdesprot-II* was 52 °C.

Expression of mdesprot-I *and* mdesprot-II *mRNA in larval tissues and during development*

RT-PCR to assess expression of *mdesprot-I* and *mdesprot-II* mRNA in larval tissues was carried out using the SuperScript™ One Step RT-PCR System with Platinum® *Taq* DNA polymerase according to the manufacturer's protocol with RNA extracted from midgut tissue, salivary glands and the remaining larval carcasses using primers designed to the ends of the coding region (*vide supra*). RT-PCR to assess expression of *mdesprot-I* and *mdesprot-II* mRNA at different times of larval development and in pupae and adults was performed as described, for the larval tissues using RNA extracted from 1st-instars (four-days old), early 2nd-instars (eight-days old), late 2nd-instars (13-days old), 3rd-instars (from puparia), pupae, and adults. Equal amounts of total RNA (1.0 μg) were used in all RT-PCRs. Different numbers of amplification cycles (i.e. 20, 25, 30, 35, 40, 45 cycles) were used to assess whether amplifications had been taken to saturation. RT-PCR with primers designed to a Hessian fly actin gene sequence (Shukle, 2000) (Act-F, 5′-ATGTGTGACGACGAAGTTGCTTTGGTT-3′; Act-R, 5′-CATAACGATGTTAGCGTACAAGTCCTTACG-3′) was used as the positive control. RT-PCR with RNA without reverse transcriptase to generate ssDNA template was used as the negative control to test for possible contaminating DNA in the isolated RNA. Additionally, PCR with genomic DNA as the template, displays a size difference in the amplicon obtained, as compared with the ssDNA template due to the introns in the genomic sequence.

Fragment isolation, cloning and sequence analysis

All the PCR fragments obtained from DNA based PCR, RT-PCR, and 5′; 3′ RACE were excised from 1% agarose gels and purified using the MinElute™ Gel Extraction kit (Qiagen, Valencia, CA) following the manufacturer's protocol. The fragments were ligated into the blunt-end vector pSTBLUE (Novagen, Madison, WI). Three independent clones of each DNA or cDNA fragment were sequenced three times using both strands by the Purdue Genomic Center. T7 and SP6 primers and synthesized oligonucleotides to determine internal sequences were used. Results for each clone were aligned and a consensus formed using the SeqWeb sequence analysis program (http://silverjack.genomics.purdue.edu) to correct for sequencing errors or errors introduced by the *Taq* polymerase. Sequence similarity and annotations were done using different BLAST programs on the National Center for Biotechnology Information (Bethesda, MD) website (http://www.ncbi.nlm.nih.gov/). Analysis for secretion signal peptides was performed using the iPSORT prediction website (http://hypothesiscreator.net/iPSORT/predict.cgi), and the signal Pv.1.1 website (http://www.cds.dtu.dk/services/signalP1). The calculated molecular weights for both proteins were obtained using the SeqWeb web-based sequence analysis program (http://silverjack.genomics.purdue.edu).

BAC library screening and fluorescence in situ *hybridization (FISH)*

Two BAC libraries of the Hessian fly were screened as by Chen *et al*. (2004) with one modification. Instead of performing the overnight hybridization in a plastic bag, it was carried out in glass tubes. The BAC libraries used in this study have been described elsewhere (Behura *et al*., 2004). FISH for genes encoding the proteases followed the protocol described by Chen *et al*. (2004) with modification. Probes for the two genes involved in the study were prepared by labelling BAC DNA with both biotin and digoxigenin-conjugates.

Phylogenetic analysis

To reveal the relationship of the Hessian fly proteases to those of other insect species, a phylogenetic tree was constructed with the amino acid sequences of midgut serine proteases from a number of Dipterans and Lepidoptera, using a pancreatic serine protease from *Homo sapiens* as the out-group. The peptide sequences were aligned with the ClustalX program, version 8.1 with 11 updates (Thompson *et al*., 1997). The phylogenetic tree was calculated by the distance/neighbour-joining method in the ClustalX program. Bootstrap values for the branches were obtained with 1000 replications.

Acknowledgements

We thank the anonymous reviewers for suggestions that greatly improved the manuscript. Technical support provided by Ting-Ting Lu and Amanda Alber during the course of this work is gratefully acknowledged. This is a joint contribution of the USDA-ARS & Purdue University. Purdue University Agricultural Experiment Journal Paper 2004–17518. This study was supported through USDA CRIS no. 3602–22000–014D.

References

Choi, I.R., Stenger, D.C., Morris, T.J. and French, R. (2000) A plant virus vector for systematic expression of foreign genes in cereals. *Plant J* **23**: 547–555.

Behura, K.S., Valicente, H.F., Rider, D.S., JrChen, M.-S., Jackson, S. and Stuart, J.J. (2004) A physically anchored genetic map and linkage to virulence reveals recombination suppression over the proximal region of Hessian fly chromosome A2. *Genetics* **167**: 343–355.

Brown, D.P., Wilkinson, H.S. and Gatehouse, J.A. (2004) Regulation of genes encoding digestive proteases in the gut of a polyphagous lepidopteran larva in response to dietary protease inhibitors. *Physiol Entomol* **29**: 278–290.

Byers, R.A. and Gallun, R.L. (1972) Ability of the Hessian fly to stunt winter wheat. I. Effect of larval feeding on elongation of leaves. *J Econ Entomol* **65**: 955–958.

Cartwright, W.B., Caldwell, R.M. and Compton, L.E. (1959) Responses of resistant and susceptible wheat to Hessian fly attack. *Agron J* **51**: 529–531.

Chen, M.-S., Fellers, J.P., Stuart, J.J., Reese, J.C. and Liu, X. (2004) A group of related cDNAs encoding secreted proteins from Hessian fly [*Mayetiola destructor* (Say)] salivary glands. *Insect Mol Biol* **13**: 101–108.

Choi, Il-Ryong, Stenger, D.C., Morris, T.J. and French, R. (2000) A plant virus vector for systematic expression of foreign genes in cereals. *Plant J* **23**: 547–555.

Cox, T.S. and Hatchett, J.H. (1994) Hessian fly-resistance gene *H26* transferred from *Triticum tauschii* to common wheat. *Crop Sci* **34**: 958–960.

Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J.A. and White, B.N. (1985) A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. *Nucleic Acids Res* **13**: 6605–6619.

El Bouhssini, M., Hatchett, J.H. and Wilde, G.E. (1998) Survival of Hessian fly (Diptera: Cecidomyiidae) larvae on wheat cultivars carrying different genes for antibiosis. *J Agri Entomol* **15**: 183–193.

El Bouhssini, M., Hatchett, J.H., Cox, T.S. and Wilde, G.E. (2001) Genotypic interaction between resistance genes in wheat and virulence genes in the Hessian fly, *Mayetiola destructor* (Diptera: Cecidomyiidae). *Bull Entomol Res* **91**: 327–331.

Gallun, R.L. (1977) Genetic basis of Hessian fly epidemics. *Ann NY Acad Sci* **287**: 223–229.

Gallun, R.L. and Langston, R. (1963) Feeding habit of Hessian fly larvae on ^{32}P-labeled resistant and susceptible wheat seedlings. *J Econ Entomol* **56**: 702–706.

Gatehouse, A.M.R., Gatehouse, J.A., Dobie, P., Kilminster, A.M. and Boulter, D. (1979) Biochemical basis of insect resistance in *Vaguna unguiculata*. *J Sci Fd Agric* **30**: 948–958.

Grover, P.B. Jr., Ross, D.R. and Shukle, R.H. (1988) Identification of digestive carbohydrases in larvae of the Hessian fly, *Mayetiola destructor* (Say) (Diptera: Cecidomyiidae). *Arch Insect Biochem Physiol* **8**: 59–72.

Gruden, K., Popovic, T., Cimerman, N., Krizaj, I. and Strukelj, B. (2003) Diverse enzymatic specificities of digestive proteases, 'intestains', enable Colorado potato beetle larvae to counteract the potato defense mechanism. *Biol Chem* **384**: 305–310.

Hatchett, J.H. and Gallun, R.L. (1970) Genetics of the ability of Hessian fly, *Mayetiola destructor*, to survive on wheat having different genes for resistance. *Ann Entomol Soc Am* **63**: 1400–1407.

Jones, E.T. (1939) Grasses of the tribe Hordeae as hosts of the Hessian fly. *J Econ Entomol* **32**: 505–510.

Kalhok, S.E., Tabak, L.M., Prosser, D.E., Brook, W., Downe, A.E.R. and White, B.N. (1993) Isolation, sequencing and characterization of two cDNA clones coding for trypsin-like enzymes from the midgut of *Aedes aegypti*. *Insect Mol Biol* **2**: 134–145.

Kollien, A.H., Waniek, P.J., Prols, F., Habedank, B. and Schaub, G.A. (2004) Cloning and Characterization of a trypsin-encoding cDNA of the human body louse *Pediculus humanus*. *Insect Mol Biol* **13**: 9–18.

Kraut, J. (1977) Serine proteases: structure and mechanisms of catalysis. *Ann Rev Biochem* **8**: 331–358.

Krem, M.M. and Cera, E.D. (2001) Molecular markers of serine protease evolution. *EMBO J* **20**: 3036–3045.

Liu, X., Fellers, J.P., Wilde, G.E., Stuart, J.J. and Chen, M.-S. (2004) Characterization of two genes expressed in the salivary glands of the Hessian fly, *Mayetiola destructor* (Say). *Insect Biochem Mol Biol* **34**: 229–237.

Lopes, A.R., Juliano, M.A., Juliano, L. and Terra, W.R. (2004) Coevolution of insect trypsins and inhibitors. *Arch Insect Biochem Physiol* **55**: 140–152.

Martin-Sanchez, J.A., Gomez-Colmenarejo, M., Del Moral, J., Sin, E., Montes, M.J., Gonzalez-Belinchon, C., Lopez-Brana, I. and Delibes, A. (2003) A new Hessian fly resistance gene (*H30*) transferred from the wild grass *Aegilops triuncialis* to hexaploid wheat. *Theor Appl Genet* **106**: 1248–1255.

Ohm, H.W., Ratcliffe, R.H., Patterson, F.L. and Cambron, S.E. (1997) Resistance to Hessian fly conditioned by genes *H19* and proposed gene *H27* of durum wheat line PI422297. *Crop Sci* **37**: 113–115.

Rahbe, Y., Ferrasson, E., Rabesona, H. and Quillien, L. (2003) Toxicity to the pea aphid *Acyrthosiphon pisum* of anti-chymotrypsin isoforms and fragments of Bowman-Birk protease inhibitors from pea seeds. *Insect Biochem Mol Biol* **33**: 299–306.

Ratcliffe, R.H., Safranski, G.G., Patterson, F.L., Ohm, H.W. and Taylor, P.L. (1994) Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. *J Econ Entomol* **87**: 1113–1121.

Ratcliffe, R.H., Cambron, S.E., Flanders, K.L., Bosque-Perez, N.A., Clement, S.L. and Ohm, H.W. (2000) Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the south eastern, Midwestern, and north western United States and virulence to resistance genes in wheat. *J Econ Entomol* **93**: 1319–1328.

Russell, V.W. and Shukle, R.H. (1997) Molecular and cytological analysis of a *mariner* transposon from Hessian fly. *J Hered* **88**: 72–76.

Shirai, Y. (2004) Influence of transgenic insecticidal crops on non-target arthropods: a review. *Jap J Ecol* **54**: 47–65.

Shukle, R.H. (2000) Molecular and cytological characterization of an actin gene from Hessian fly (Diptera: Cecidomyiidae). *Ann Entomol Soc Am* **93**: 1164–1172.

Shukle, R.H. (2003) *Hessian fly* Mayetiola destructor *(Say) (Diptera: Cecidomyiidae).* Encyclopedia of Entomology (Capinera, J.L., ed.), pp. 27–31.

Shukle, R.H., Murdock, L.L. and Gallun, R.L. (1985) Identification and partial characterization of a major gut proteinase from larvae of the Hessian fly, *Mayetiola destructor* (Say) (Diptera: Cecidomyiidae). *Insect Biochem* **15**: 93–101.

Stuart, J.J. and Hatchett, J.H. (1988) Cytogenetics of the Hessian fly. I. Karyotype analysis and polytene chromosome correlations. *J Hered* **79**: 184–189.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**: 4875–4882.

Volpicella, M., Ceci, L.R., Cordewener, J., America, T., Gallerani, R., Bode, W., Jongsma, M.A. and Beekwilder, J. (2003) Properties of purified gut trypsin from *Helicoverpa zea*, adapted to proteinase inhibitors. *Eur J Biochem* **270**: 10–19.

Williams, C.E., Collier, C.C., Sardesai, N., Ohm, H.W. and Cambron, S.E. (2003) Phenotypic assessment and mapped markers for *H31*, a new wheat gene conferring resistance to Hessian fly (*Diptera: Cecidomyiidae*). *Theor Appl Genet* **107**: 1516–1523.

Yoshiyama, M. and Shukle, R.H. (2004) Molecular cloning and characterization of a glutathione *S*-transferase gene from Hessian fly (Diptera: Cecidomyiidae). *Ann Entomol Soc Am* **97**: 1285–1293.

Zantoko, L. and Shukle, R.H. (1997) Genetics of virulence in the Hessian fly to resistance gene *H13* in wheat. *J Hered* **88**: 120–123.

Zavala, J.A., Patankar, A.G., Gase, K., Hui, D. and Baldwin, I.T. (2004) Manipulation of endogenous trypsin proteinase inhibitor production in *Nicotiana attenuate* demonstrates their function as anti-herbivore defences. *Plant Physiol* **134**: 1181–1190.

Zhu, Y.-C., Kramer, K.J., Dowdy, A.K. and Baker, J.E. (2000) Trypsinogen-like cDNAs and quantitative analysis of mRNA levels from the Indianmeal moth, *Plodia interpunctella*. *Insect Biochem Mol Biol* **30**: 1027–1035.